

## SMP: DNA

## Project: Towards a SNP Based Haplotype Map for the Rat Genome

Norbert Hübner – Max-Delbrück-Center for Molecular Medicine (MDC), Berlin-Buch  
nhuebner@mdc-berlin.de

## Introduction

Determining the molecular basis of natural phenotypic variation, including inter-individual susceptibility to common diseases, is a central challenge of post-genome genetics.

The rat provides the most relevant models of common multifactorial human disease, and it is by far the leading model species in pharmacology and toxicology. It has been the major model for physiological investigation, providing a body of data on pathophysiology, including detailed mechanistic, biochemical and metabolic characterisation that cannot be replaced by other models. Decades of exquisite phenotyping and detailed analysis of crosses of inbred rats have resulted in initial localization of hundreds of loci involved in complex disease and quantitative phenotypes, but with very few eventual gene identifications to date. A clear understanding of the origin and structure of genetic variation in the rat will provide a key missing piece of this puzzle. To fully realize the power of the recent rat genome sequence (1), we propose to initiate the complete genetic dissection of the ancestral segments making up the most commonly used inbred lines. The proposed SNP based haplotype map provides a valuable tool for functional genomics, specifically by focussing positional cloning of QTLs through i) the reduction of regions obtained through linkage analysis via identification of segments shared by the strains used for the cross; ii) the selection of ideal strain combinations for further reduction of critical regions through simple intercross/backcross experiments; iii) the use of correlation between phenotype and ancestral sequence origin across many inbred strains to identify very short genomic regions most likely to harbour responsible genes.

The aim of this project is to construct a detailed SNP based haplotype map for the rat genomic regions that are syntenic to human chromosome 21. These genomic regions were chosen, since there is detailed knowledge about the genomic landscape of human chr. 21 that encompasses several QTLs for hypertension, insulin levels, glucose levels and is particularly important for its role in human trisomy 21. Together with our collaborators at the Max-Planck-Institute for Molecular Genetics, Berlin (Richard Reinhardt) and at the Institute for Molecular Biology, Jena (Matthias Platzer) we will identify one SNP every 30kb within the relevant regions of rat chr. 11 and chr. 20 that represent the syntenic regions to human chr. 21. We will then proceed to genotype these SNPs across 50 most commonly used inbred rat strains to identify the genomic architecture in these strains and identify those segments that are shared identity by descent from the most recent common ancestor.

## Results/Project Status

Presently, we are identifying candidate SNP that will be amenable for high-density genotyping across the selected genomic regions. Currently, more than 10,000 candidate SNPs have been identified from genomic and cDNA libraries constructed from four distantly related rat strains. We hope that once a detailed map describing the genetic architecture in these 50 rat strains has been established that this data can be used to relate shared IBD genomic regions to particular physiological traits. This may help prioritise regions that may be particularly likely to harbor genetic variants that underlie complex disease phenotypes.

In a pilot study we have recently investigated the overall effect of polymorphisms on gene expression levels we compared the SNP frequency across the genome with the SNP frequency in expression QTL (eQTL) genes. We found a

highly significant enrichment of SNPs in the cis-regulated eQTL genes compared with either the trans-regulated eQTL genes or the observed rate across the genome (2). Cis-acting eQTLs are of particular interest as positional candidate genes for QTLs. We applied a comparative mapping strategy to explore the applicability of the detected cis-acting eQTLs to human hypertension. By forming a robust dataset of cis-acting eQTL genes with a false discovery rate <5% that were contained within rat blood pressure related QTLs we identified the human orthologs and compared with the location of previously mapped human blood pressure QTLs. This analysis defined a set of 73 attractive candidate genes for testing in human hypertension data sets (3).

## Further activities

The efforts to investigate the haplotype structure within the rat genome that were initiated with the present proposal lead to two funded EU 6 initiatives within the 6<sup>th</sup> framework program that is expanding these efforts here exemplified for the rat syntenic regions of human chr. 21 to the entire rat genome. These EU based consortia are coordinated by NH at the MDC and involve participation of the MPI and IMB.

## Outlook

The analysis of gene function based on the examination of genetic variation is a very promising strategy for the identification of responsible genes underlying genetic disorders in complex phenotypes. The exploitation of sequence variation in the rat will help to understand more about the molecular processes that play a key role in the molecular evolution of complex physiological and pathophysiological phenotypes that may be shared in common with similar disorders in humans.



**Fig 1:** The rat on a chip – for high throughput genotyping and expression profiling.

Lit.: 1. Rat Genome Sequencing Project Consortium. (2004). Genome Sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature*. 2004. 428, 493-521. 2. Hübner N, Wallace CA, Zimdahl H, et al. Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat Genet*. 2005. 37, 243-253. 3. Zimdahl H, Nyakatura G, Brandt P, Schulz H, Hummel O, Fartmann B, Brett D, Droege M, Monti J, Lee YA, Sun Y, Zhao S, Winter EE, Ponting CP, Chen Y, Kasprzyk A, Birney E, Ganten D, Hübner N. A SNP map of the rat genome generated from cDNA sequences. *Science*. 2004. 303, 807.